

STATISTICAL METHODS FOR THE EVALUATION OF TOURISM SERVICE

2021/2022

Teacher: Mariangela Zenga

Creado por:

Miguel Díaz-Plaza Cabrera

Barbara Semedo Alves Goncalves



[ConjuntoDatos1] C:\Users\migue\AppData\Local\Temp\Temp1_Exam ERASMUS.zip\Dataset_1.sav

The dataset Dataset_1.sav contains the answers of a ISTAT (2011) survey on discrimination based on gender, sexual orientation, ethnicity of 2011. The data set contains the following variables:

1) Define and correctly describe the variables GENDER, AGE, EDUC, COND and NCOMP giving the graphical representation and correctly commenting indexes of position and variability.

Definition and description of the variable 'Gender' and 'Work condition'

GENDER

Is a numeric and nominal variable with 2 categories:

1 is for male

2 is for female

CONDIZ (WORK CONDITION)

Is a numeric and nominal variable with 6 categories:

1 is for employed

2 is for seeking for a job

4 is for housewife

5 is for student

6 is for retired

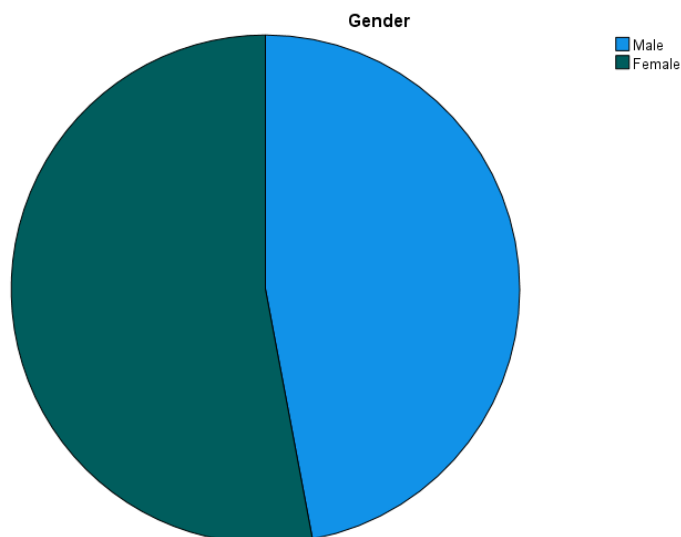
8 is for other condition

Estadísticos

| | | Gender | Work condition |
|------|----------|--------|----------------|
| N | Válido | 304 | 304 |
| | Perdidos | 0 | 0 |
| Moda | | 2 | 1 |

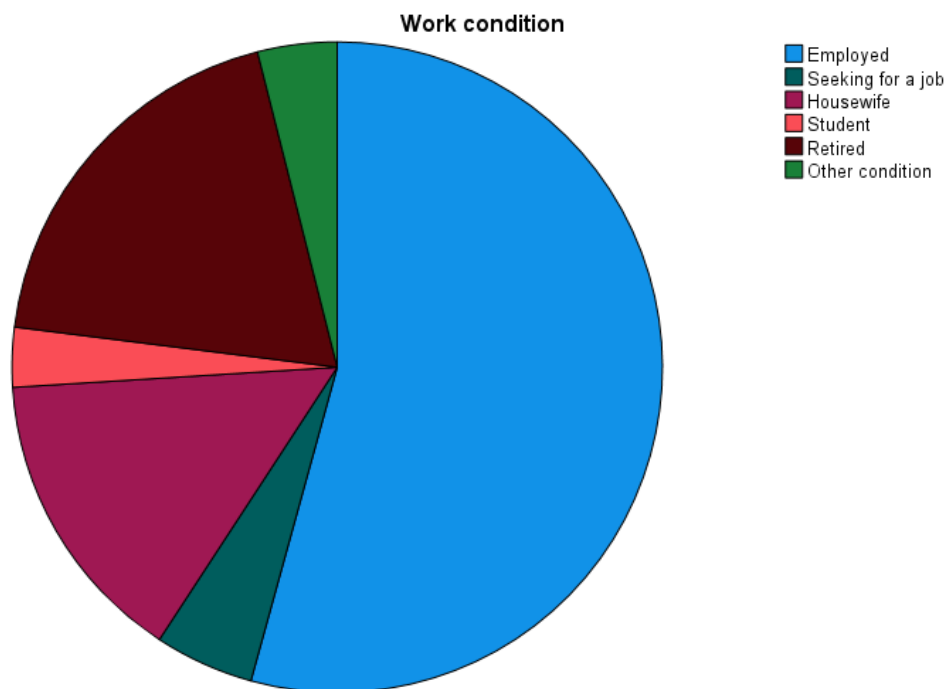
- The mode is the value that appears most frequently in a data set. A set of data may have one mode, more than one mode, or no mode at all. In this data set, number 1 corresponds to 'Male' and the number 2 corresponds to 'Female'. As we can see in the table, in the category 'Gender' the mode is 2, Female, and in the category 'Work condition' the mode is 1, Male.

| | | Gender | | | |
|--------|--------|------------|------------|-------------------|----------------------|
| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
| Válido | Male | 143 | 47,0 | 47,0 | 47,0 |
| | Female | 161 | 53,0 | 53,0 | 100,0 |
| | Total | 304 | 100,0 | 100,0 | |



- The variable Gender is a numeric and nominal variable with two categories, female, and male. In total, male and female, there are 304 people in this study. We can conclude that in terms of frequency, the female gender is more numerous than the male gender 161 and 143 respectively, with 18 more women than men. In percentage terms, these results are evident, where we can see 53% of the female gender and 47% of the male gender.

| | | Work condition | | | |
|--------|-------------------|----------------|------------|-------------------|----------------------|
| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
| Válido | Employed | 165 | 54,3 | 54,3 | 54,3 |
| | Seeking for a job | 15 | 4,9 | 4,9 | 59,2 |
| | Housewife | 45 | 14,8 | 14,8 | 74,0 |
| | Student | 9 | 3,0 | 3,0 | 77,0 |
| | Retired | 58 | 19,1 | 19,1 | 96,1 |
| | Other condition | 12 | 3,9 | 3,9 | 100,0 |
| | Total | 304 | 100,0 | 100,0 | |



- The variable 'Work Condition' is a numeric and nominal variable with six categories, which are, employed, seeking for a job, housewife, student, retired and other condition. The total number of people used for this sample is 304.

We can see that the category with the most weight is the 'Employed' category with 165 in terms of frequency, that is, 165 people out of 304 are employed. In percentage terms, 54% of the total sample is employed, more than half.

The categories 'Retired' and 'Housewife' have a median contribution, presenting in terms of frequency close to 58 and 45 respectively. In percentage terms, they present 19% and 14% respectively.

The categories that have less weight in the sample are 'seeking for a job', 'Other condition' and 'Student', in terms of frequency, 15, 12 and 9 respectively. In percentage terms, they present minimum values, 4.9%, 3.9% and 3% respectively.

- We can confirm in the circular graph that, visually, we can see that the most significant portion is blue, which corresponds to 'Employed' and the least significant, pink, corresponds to 'Student' category.

Indexes of position and variability

As they are nominal variables, we could do a bar chart, but we decided to do a circular graph, so it's possible to have a better vision of all the categories. The central tendency is study only by the mode and the variability is studied by the index qualitative variation or variation ratio.

Definition and description of the variable 'Age' and 'Level of Education'

AGE

Is a numeric and ordinal variable with 5 categories:

1 is for 18-34

2 is for 35-44

3 is for 45-54

4 is for 55-64

5 is for 65--74

EDUC (LEVEL OF EDUCATION)

Is a numeric and ordinal variable with 4 categories:

1 is for Primary school

2 is for Secondary school

3 is for High school

4 is for University or more

Estadísticos

| | | Age | Level of Education |
|-------------|----------|------|--------------------|
| N | Válido | 304 | 304 |
| | Perdidos | 0 | 0 |
| Mediana | | 3,00 | 3,0000 |
| Moda | | 2 | 3,00 |
| Rango | | 4 | 3,00 |
| Percentiles | 25 | 2,00 | 2,0000 |
| | 50 | 3,00 | 3,0000 |
| | 75 | 4,00 | 3,0000 |

- The median is the value that occupies the central position of the data set if we order them from highest to lowest or vice versa. As we can see the median in both categories are 3 which is equivalent to the range of age class between 45-54. This is the central class range, existing 2 more before, '18-34' and '35-44', and 2 range classes after which are '55-64' and '65-74'.

- The mode is the value within the data set that is repeated the most, the one with the highest frequency. In the table we can observe that in the variable 'Age' the mode is 2, which means the class range '35-44'. This means that there are more people aged between this range in the sample.

- The range establishes the proximity of the data in the set. It is calculated by subtracting the smallest data from the largest data. For this, we must subtract the minimum from the maximum. In the variable 'age' the range (interval), is 4 resulting from subtraction $5-1=4$. The number 4 is equivalent to the range class 55-64. The

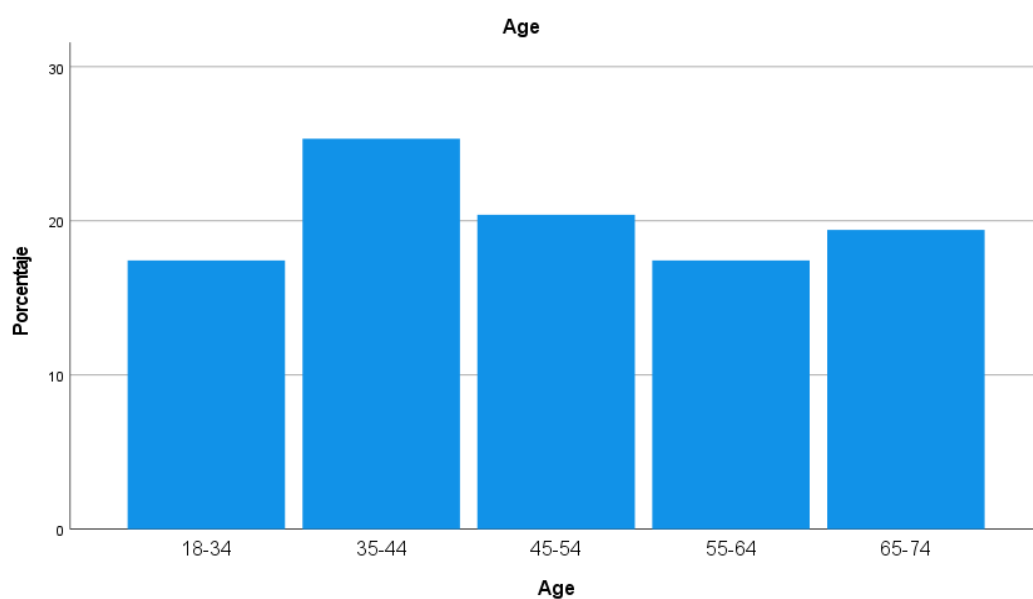
interquartile range is the difference between Q3-Q1, or in other words, percentile 75 – percentile 25. IQR for Age is 2.00 and for Level of Education 1.00. In the variable 'level of education', the range is 3 resulting from the subtraction 4-1=3. The number 3 is equivalent to the class '45-54'.

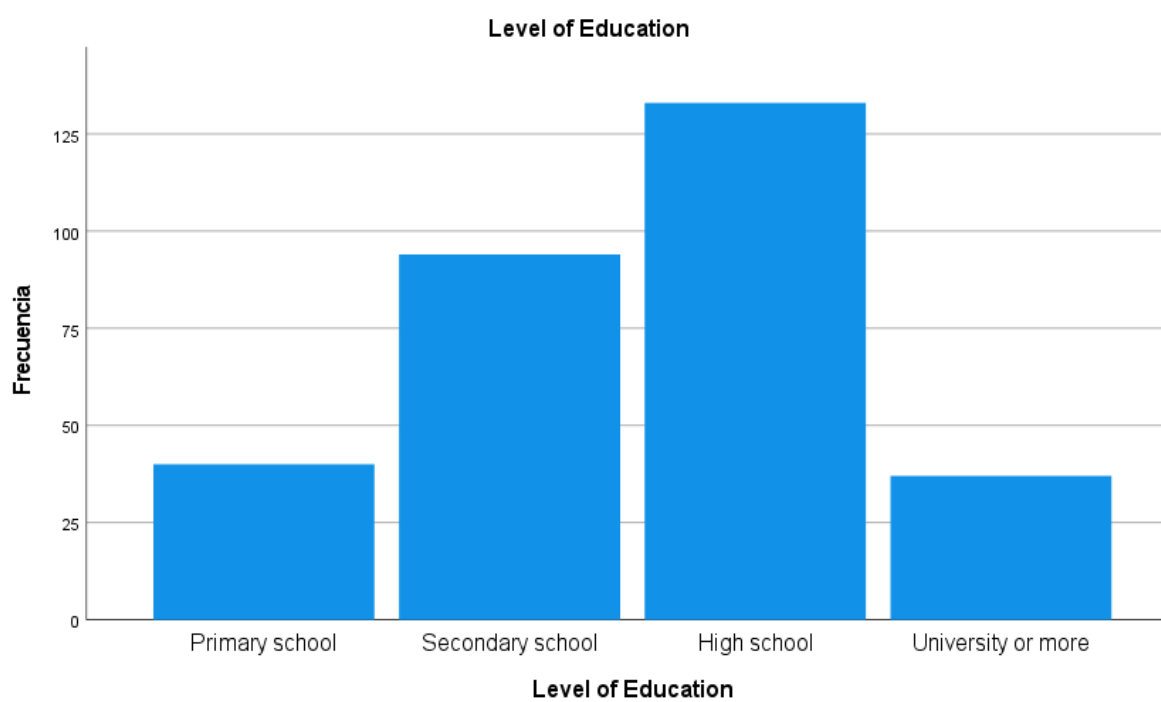
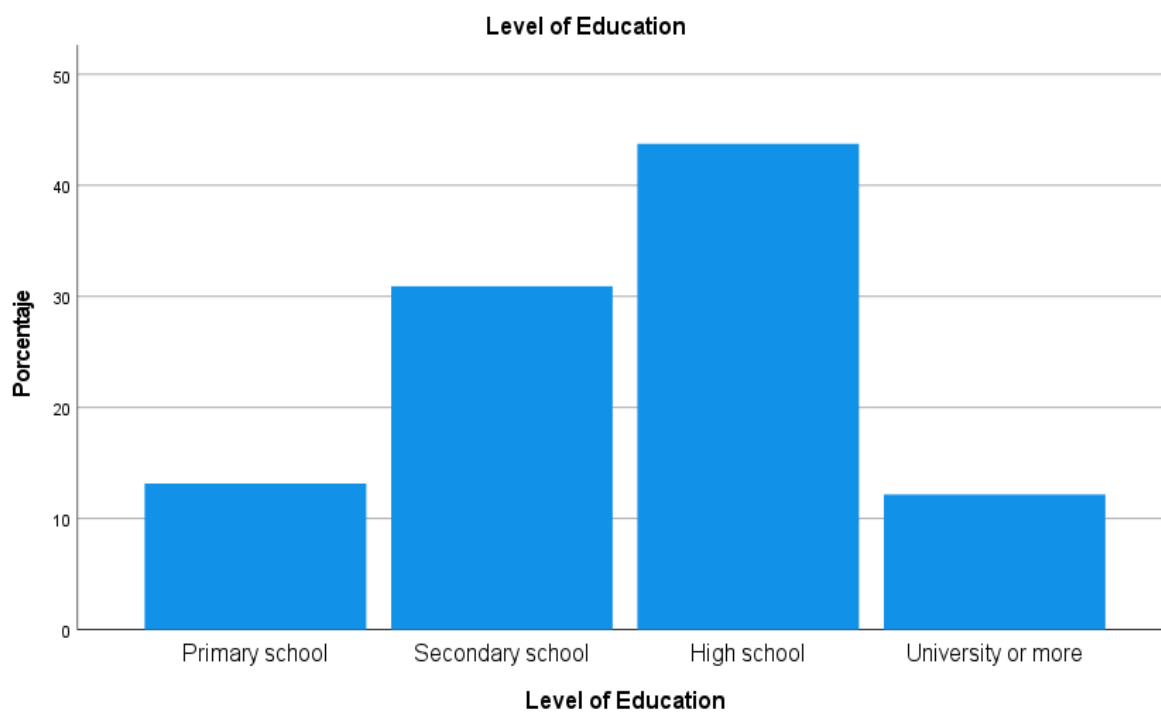
Age

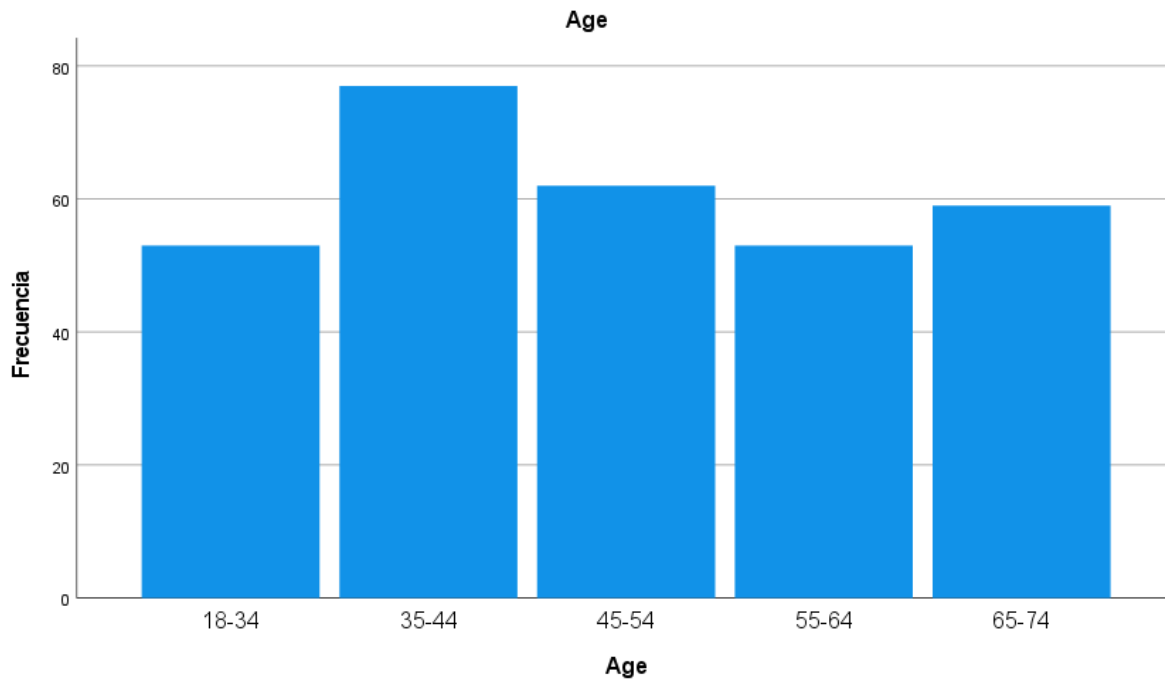
| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|--------|-------|------------|------------|-------------------|----------------------|
| Válido | 18-34 | 53 | 17,4 | 17,4 | 17,4 |
| | 35-44 | 77 | 25,3 | 25,3 | 42,8 |
| | 45-54 | 62 | 20,4 | 20,4 | 63,2 |
| | 55-64 | 53 | 17,4 | 17,4 | 80,6 |
| | 65-74 | 59 | 19,4 | 19,4 | 100,0 |
| | Total | 304 | 100,0 | 100,0 | |

Level of Education

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|--------|--------------------|------------|------------|-------------------|----------------------|
| Válido | Primary school | 40 | 13,2 | 13,2 | 13,2 |
| | Secondary school | 94 | 30,9 | 30,9 | 44,1 |
| | High school | 133 | 43,8 | 43,8 | 87,8 |
| | University or more | 37 | 12,2 | 12,2 | 100,0 |
| | Total | 304 | 100,0 | 100,0 | |







Indexes of position and variability

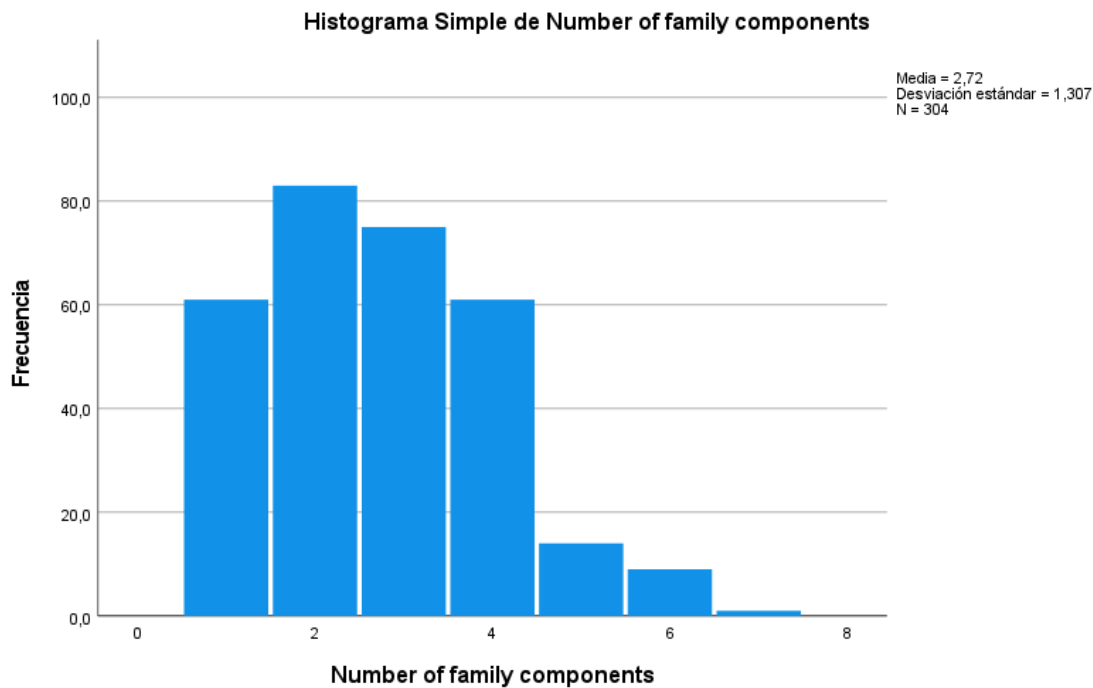
- As they are ordinal variables, we can do a bar charts (not histograms because the date is not continuous), the central tendency summarized by median and mode (not mean), and the variability by range and interquartile range (not standard deviation)

Definition and description of the variable 'Number of family components'

NCOMP is a scale variable

Estadísticos descriptivos

| | N | Mínimo | Máximo | Media | Desviación estándar |
|-----------------------------|-----|--------|--------|-------|---------------------|
| Number of family components | 304 | 1 | 7 | 2,72 | 1,307 |
| N válido (por lista) | 304 | | | | |



2) Choose and analyze one variable among the V1-V5 through the frequency table and the appropriate plot. Adequately comment on the results.

| | |
|----|---|
| V1 | Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience |
| V2 | Are women better or worse: on earnings earned for the same job? |
| V3 | Are women better or worse: on career opportunities / promotion? |
| V4 | Are women better or worse: on job stability? |
| V5 | Are women better or worse: on adequate consideration by the employer and co-workers? |

Are ordinal variables from 1 to 3 being

1 Worse

2 No difference

3 Better

We have chosen V2. Are women better or worse: on earnings earned for the same job?

We considered it as a controversial question, and we want to discover if the inequalities have been reduced in the recent days or not.

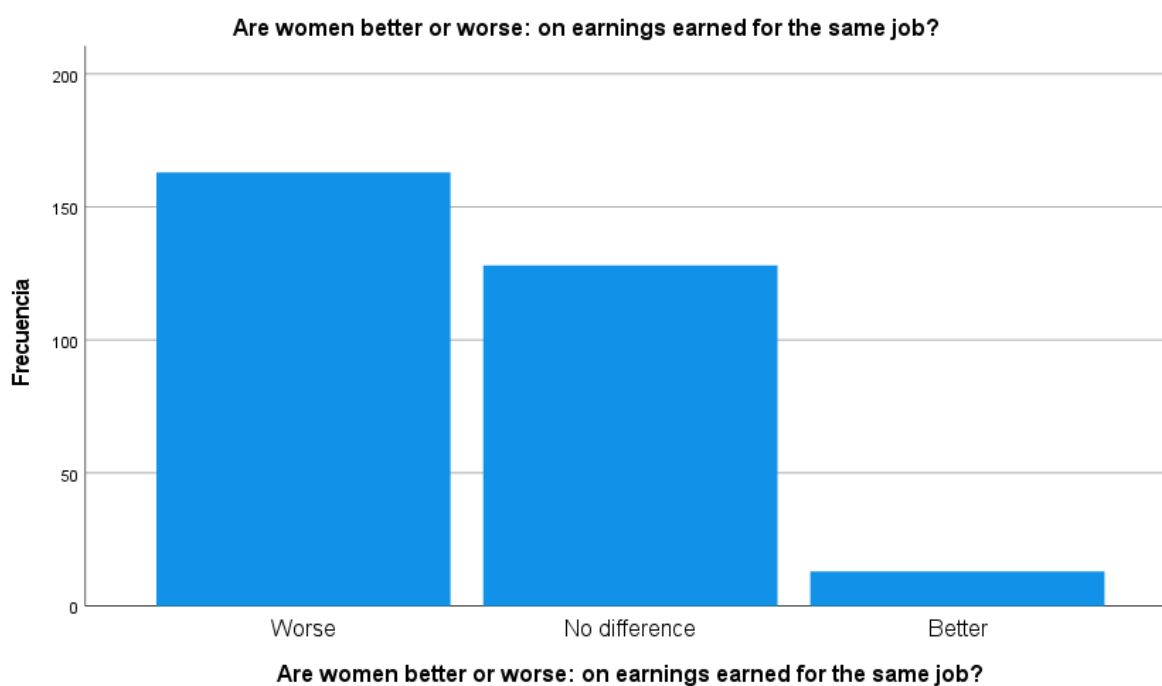
Estadísticos

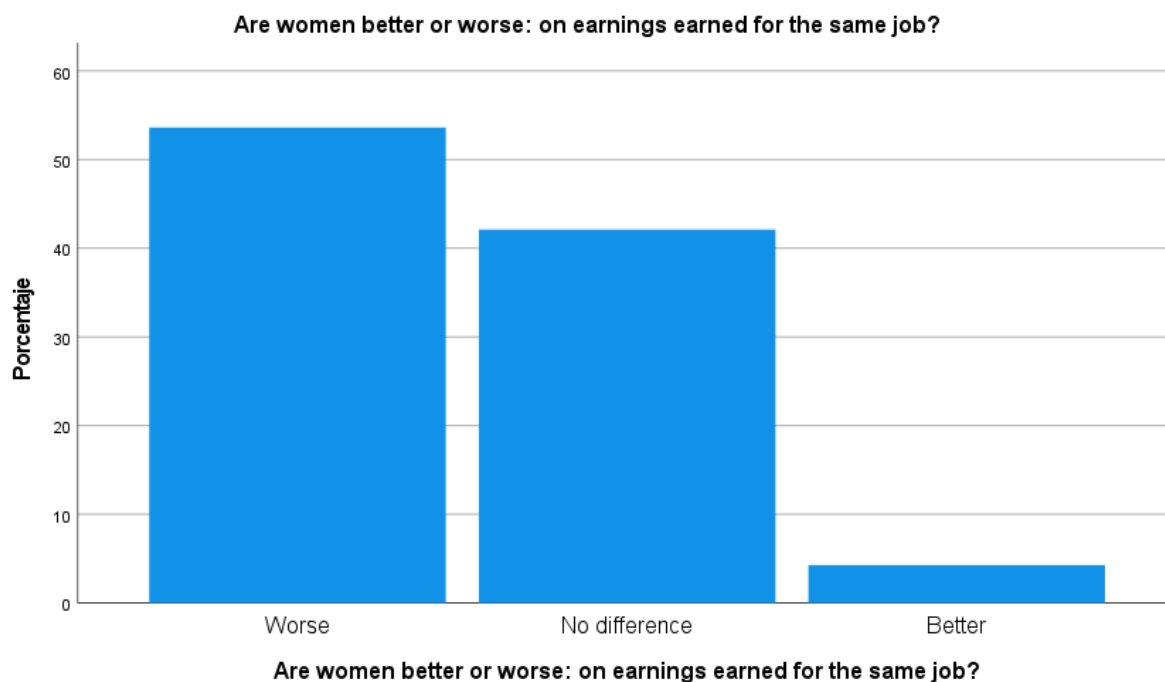
Are women better or worse: on earnings earned for the same job?

| | | |
|---|----------|-----|
| N | Válido | 304 |
| | Perdidos | 0 |

Are women better or worse: on earnings earned for the same job?

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|--------|---------------|------------|------------|-------------------|----------------------|
| Válido | Worse | 163 | 53,6 | 53,6 | 53,6 |
| | No difference | 128 | 42,1 | 42,1 | 95,7 |
| | Better | 13 | 4,3 | 4,3 | 100,0 |
| | Total | 304 | 100,0 | 100,0 | |

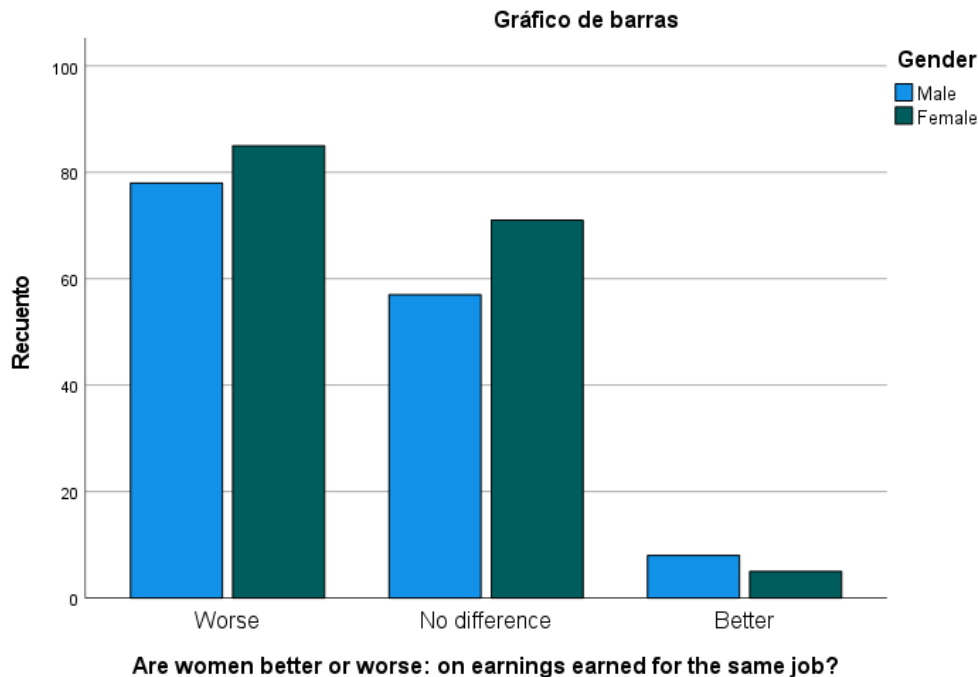




We can see that half more of the people on the survey admit that women are worse on earnings earned for the same job with 53.6% and only 4.3% of the total say it is better. We take in account it is a little bit ambiguous question because it depends on the type of job, but generally talking, these are the responses.

Tabla cruzada Are women better or worse: on earnings earned for the same job?*Gender

| | | | Gender | | Total |
|---|--------------------|--------------------|--------|--------|--------|
| | | | Male | Female | |
| Are women better or worse: on earnings earned for the same job? | Worse | Recuento | 78 | 85 | 163 |
| | | % dentro de Gender | 54,5% | 52,8% | 53,6% |
| | No difference | Recuento | 57 | 71 | 128 |
| | | % dentro de Gender | 39,9% | 44,1% | 42,1% |
| | Better | Recuento | 8 | 5 | 13 |
| | | % dentro de Gender | 5,6% | 3,1% | 4,3% |
| Total | Recuento | | 143 | 161 | 304 |
| | % dentro de Gender | | 100,0% | 100,0% | 100,0% |



As in our dataset we don't have the perception and the expectation at this question, we decided to a cross table and graphics comparing the results of V2 based on the gender. So, we can analyze that female tends to see they are worse or no difference and male think more they are better compared to the females that say are better.

3) Evaluate the reliability of the variables test V1-V5. Comment out properly the result.

Resumen de procesamiento de casos

| | | N | % |
|-------|-----------------------|-----|-------|
| Casos | Válido | 304 | 100,0 |
| | Excluido ^a | 0 | ,0 |
| | Total | 304 | 100,0 |

a. La eliminación por lista se basa en todas las variables del procedimiento.

There is no missing data in this dataset.

Estadísticas de fiabilidad

| Alfa de Cronbach basada en elementos estandarizados | | |
|---|------------------|----------------|
| Alfa de Cronbach | Alfa de Cronbach | N de elementos |
| ,826 | ,827 | 5 |

As we can see, 82.6% of the data variability is explained by these 5 variables. It is a good measure, but not excellent (no more than 85%)

Estadísticas de elemento

| | Media | Desviación estándar | N |
|---|-------|---------------------|-----|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | 1,56 | ,691 | 304 |
| Are women better or worse: on earnings earned for the same job? | 1,51 | ,580 | 304 |
| Are women better or worse: on career opportunities / promotion? | 1,55 | ,702 | 304 |
| Are women better or worse: on job stability? | 1,51 | ,608 | 304 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | 1,70 | ,643 | 304 |

We observe that the variable with the highest score is Are women better or worse: on adequate consideration by the employer and co-workers? with 1.7 out of 3, which means that is the better task women are better for the people on the survey. The variable with the lower score is Are women better or worse: on earnings earned for the same job? and Are women better or worse: on job stability? with 1.51, which is normal because they are the two more controversial problems women have in order to equality.

Matriz de covarianzas entre elementos

| | Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | Are women better or worse: on earnings earned for the same job? | Are women better or worse: on career opportunities / promotion? | Are women better or worse: on job stability? |
|---|---|---|---|--|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | ,478 | ,222 | ,251 | ,222 |
| Are women better or worse: on earnings earned for the same job? | ,222 | ,337 | ,203 | ,185 |
| Are women better or worse: on career opportunities / promotion? | ,251 | ,203 | ,493 | ,225 |
| Are women better or worse: on job stability? | ,222 | ,185 | ,225 | ,369 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | ,177 | ,141 | ,216 | ,192 |

Matriz de covarianzas entre elementos

Are women better or worse: on adequate consideration by the employer and co-workers?

| | |
|---|------|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | ,177 |
| Are women better or worse: on earnings earned for the same job? | ,141 |
| Are women better or worse: on career opportunities / promotion? | ,216 |
| Are women better or worse: on job stability? | ,192 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | ,414 |

In the covariance matrix we can observe the relations between variables and on the diagonal of this matrix are the variance of each variable.

Estadísticas de elemento de resumen

| | Media | Mínimo | Máximo | Rango | Máximo / Mínimo | Varianza |
|-----------------------------|-------|--------|--------|-------|-----------------|----------|
| Medias de elemento | 1,567 | 1,507 | 1,704 | ,197 | 1,131 | ,006 |
| Varianzas de elemento | ,418 | ,337 | ,493 | ,156 | 1,464 | ,005 |
| Covarianzas entre elementos | ,203 | ,141 | ,251 | ,110 | 1,786 | ,001 |

Estadísticas de elemento de resumen

N de elementos

| | |
|-----------------------------|---|
| Medias de elemento | 5 |
| Varianzas de elemento | 5 |
| Covarianzas entre elementos | 5 |

Estadísticas de total de elemento

| | Media de escala si el elemento se ha suprimido | Varianza de escala si el elemento se ha suprimido | Correlación total de elementos corregida | Correlación múltiple al cuadrado |
|---|--|--|--|--|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | 6,27 | 3,935 | ,636 | ,425 |
| Are women better or worse: on earnings earned for the same job? | 6,33 | 4,320 | ,622 | ,409 |
| Are women better or worse: on career opportunities / promotion? | 6,29 | 3,875 | ,648 | ,420 |
| Are women better or worse: on job stability? | 6,32 | 4,140 | ,666 | ,445 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | 6,13 | 4,293 | ,544 | ,316 |

Estadísticas de total de elemento

| | Alfa de Cronbach si el elemento se ha suprimido |
|---|---|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | ,787 |
| Are women better or worse: on earnings earned for the same job? | ,792 |
| Are women better or worse: on career opportunities / promotion? | ,784 |
| Are women better or worse: on job stability? | ,779 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | ,813 |

This table indicates how it change the Alfa Cronbach is we delete each variable and maintain the rest. We observe that the highest is 81.3% if we delete Are women better or worse: on adequate consideration by the employer and co-workers? which means that is the least important variable, because if we omitted it, the variability decreases only from 82.6% to 81.3%, which it is quite surprising, because no one increase the Alfa Cronbach, which means all variables are important. On the other hand, we have 77.9% if we delete Are women better or worse: on job stability? which means we can say is the most important variable according to data variability, without it we lose variability explained.

4) The V1-V5 variables are on a Likert scale. After having defined its characteristics, build a variable that measures WOMEN STATUS as the sum of the variables from V1 to V5

As we said before, our dataset we don't have the perception and the expectation at this question, we decided to a cross table and graphics comparing the results of V1 to V5 based on the gender.

Resumen de procesamiento de casos

| | Incluido | | Casos Excluido | | Total | |
|--|----------|------------|----------------|------------|-------|------------|
| | N | Porcentaje | N | Porcentaje | N | Porcentaje |
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience * Gender | 304 | 100,0% | 0 | 0,0% | 304 | 100,0% |
| Are women better or worse: on earnings earned for the same job? * Gender | 304 | 100,0% | 0 | 0,0% | 304 | 100,0% |
| Are women better or worse: on career opportunities / promotion? * Gender | 304 | 100,0% | 0 | 0,0% | 304 | 100,0% |
| Are women better or worse: on job stability? * Gender | 304 | 100,0% | 0 | 0,0% | 304 | 100,0% |
| Are women better or worse: on adequate consideration by the employer and co-workers? * Gender | 304 | 100,0% | 0 | 0,0% | 304 | 100,0% |

Informe

| Gender | | Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | Are women better or worse: on earnings earned for the same job? | Are women better or worse: on career opportunities / promotion? | Are women better or worse: on job stability? |
|--------|------------------|---|---|---|--|
| Male | N | 143 | 143 | 143 | 143 |
| | Media | 1,65 | 1,51 | 1,66 | 1,62 |
| | Mediana | 1,00 | 1,00 | 1,00 | 2,00 |
| | Desv. Desviación | ,734 | ,604 | ,752 | ,638 |
| Female | N | 161 | 161 | 161 | 161 |
| | Media | 1,48 | 1,50 | 1,45 | 1,42 |
| | Mediana | 1,00 | 1,00 | 1,00 | 1,00 |
| | Desv. Desviación | ,643 | ,560 | ,642 | ,566 |
| Total | N | 304 | 304 | 304 | 304 |
| | Media | 1,56 | 1,51 | 1,55 | 1,51 |
| | Mediana | 1,00 | 1,00 | 1,00 | 1,00 |
| | Desv. Desviación | ,691 | ,580 | ,702 | ,608 |

Informe

Are women better or worse: on adequate consideration by the employer and co-workers?

| Gender | | |
|--------|------------------|------|
| Male | N | 143 |
| | Media | 1,77 |
| | Mediana | 2,00 |
| | Desv. Desviación | ,679 |
| Female | N | 161 |
| | Media | 1,65 |
| | Mediana | 2,00 |
| | Desv. Desviación | ,606 |
| Total | N | 304 |
| | Media | 1,70 |
| | Mediana | 2,00 |
| | Desv. Desviación | ,643 |

Informe

Media

| Gender | Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | Are women better or worse: on earnings earned for the same job? | Are women better or worse: on career opportunities / promotion? | Are women better or worse: on job stability? | Are women better or worse: on adequate consideration by the employer and co-workers? |
|--------|---|---|---|--|--|
| Male | 1,65 | 1,51 | 1,66 | 1,62 | 1,77 |
| Female | 1,48 | 1,50 | 1,45 | 1,42 | 1,65 |
| Total | 1,56 | 1,51 | 1,55 | 1,51 | 1,70 |

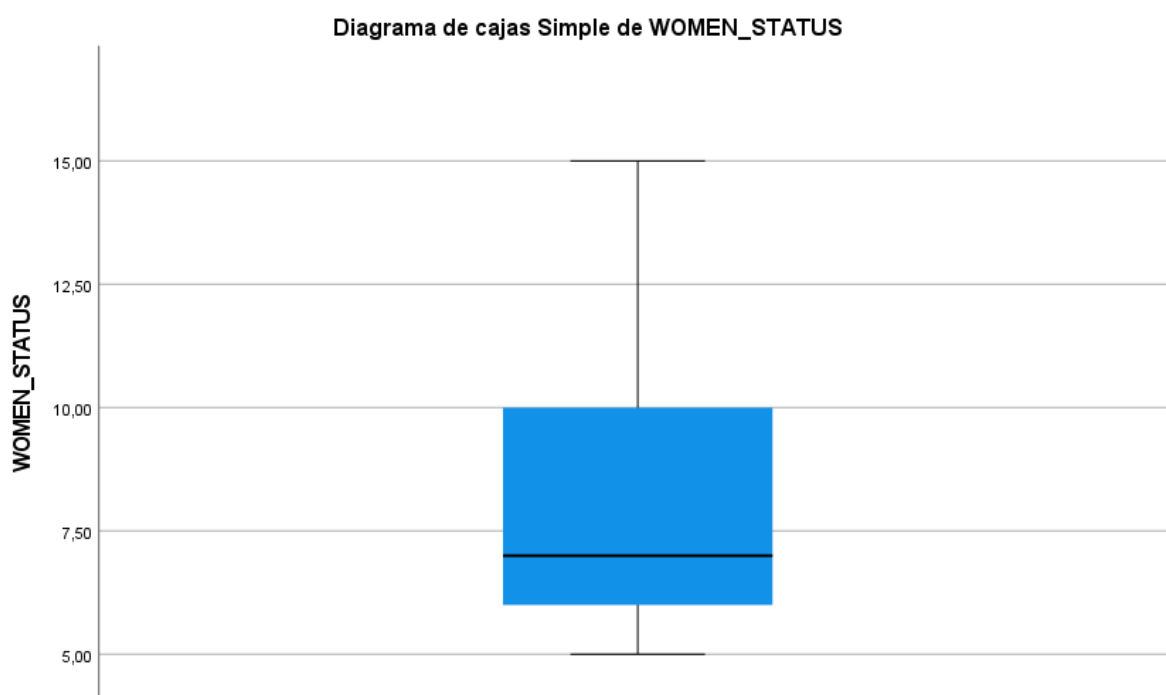
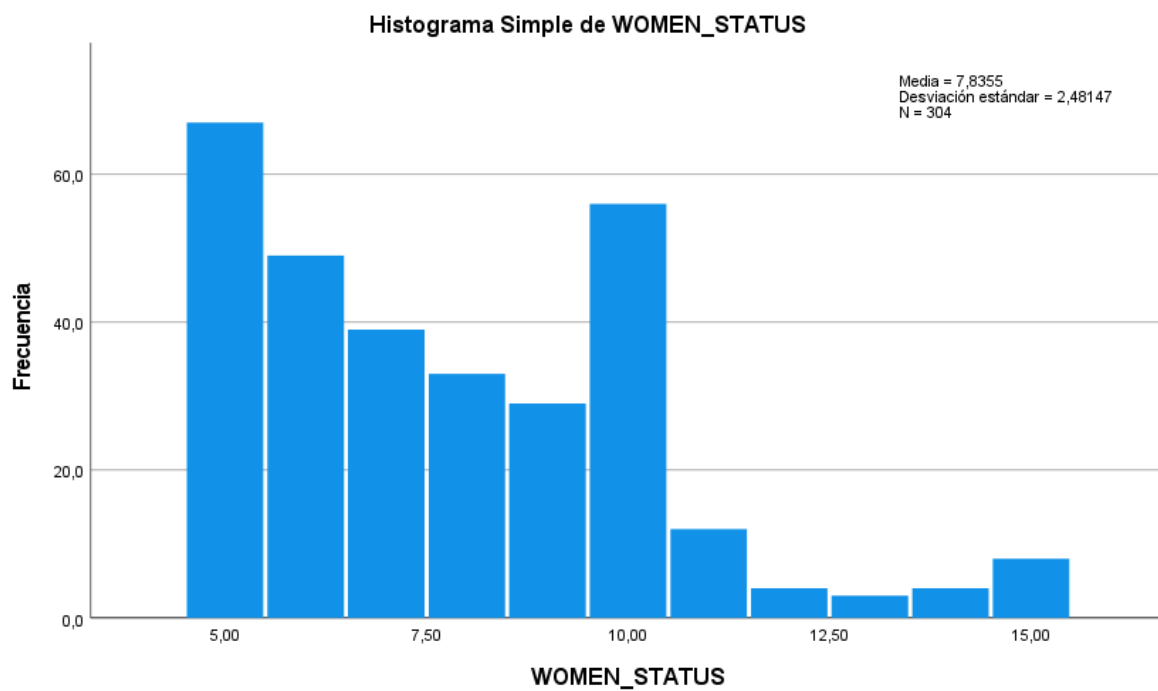
| Informe | | | | | |
|---------------------------|---|---|---|--|--|
| Media | Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | Are women better or worse: on earnings earned for the same job? | Are women better or worse: on career opportunities / promotion? | Are women better or worse: on job stability? | Are women better or worse: on adequate consideration by the employer and co-workers? |
| Gender | | | | | |
| Male | 1,65 | 1,51 | 1,66 | 1,62 | 1,77 |
| Female | 1,48 | 1,50 | 1,45 | 1,42 | 1,65 |
| Total | 1,56 | 1,51 | 1,55 | 1,51 | 1,70 |
| DIFFERENCES (MALE-FEMALE) | 0,17 | 0,01 | 0,20 | 0,19 | 0,12 |
| MEAN DIFFERENCES | 0,14 | | | | |

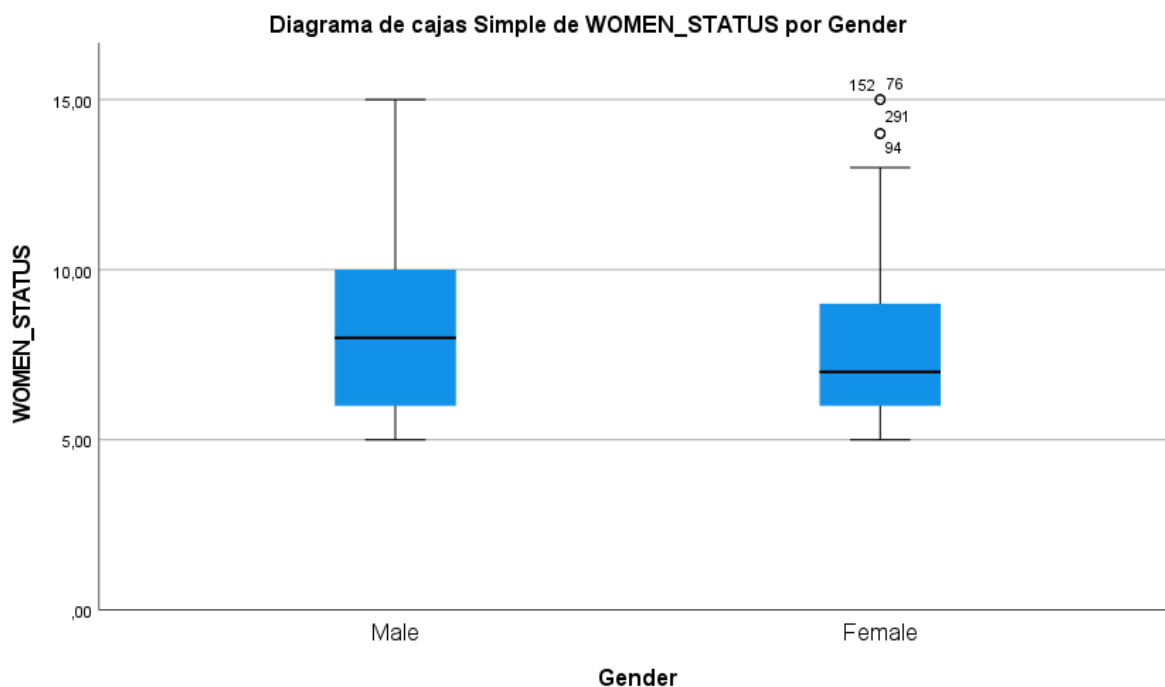
So, we have translated the table to Excel and calculate the difference between male-female for each variable and the mean of all 5. We get that male overall mean score is 0.14 points over female, and the top one with bigger difference is V3: Are women better or worse: on career opportunities / promotion? with 0.2. All 5 variables have in common that male mean score is over female mean score.

5) Describe the variable “WOMEN STATUS” according to the statistical techniques you know (Plot, Mean, Median, Variability).

Estadísticos descriptivos

| | N | Mínimo | Máximo | Media | Desviación estándar |
|----------------------|-----|--------|--------|--------|---------------------|
| WOMEN_STATUS | 304 | 5,00 | 15,00 | 7,8355 | 2,48147 |
| N válido (por lista) | 304 | | | | |





6) Apply the PCA on V1-V5 and comment out the results.

The central idea is based on the concept of the proportion of total variance that is accounted for by each of the new variables.

PCA transforms the set of the correlated variables (x_1, x_2, \dots, x_p) to a set of uncorrelated variables (y_1, y_2, \dots, y_p) called principal components, in such a way that y_1 explains the maximum possible of the total variable, y_2 the maximum of the remaining variance, and so on.

In this particular case, our 5 variables (V1,V2,V3,V4,V5) are quantitative variables, which is required for a PCA Analysis.

In SPSS, we go to Analyze, Reduction of Dimensions and Factor. Then, we pass our 5 variables. On descriptive, we chose coefficient matrix and calculate the determinant. On extraction, we check there is the PCA (Principal Components Analysis) and we are going to analyze correlation matrix and the standardized variables (that means deduct the mean and divided by standard deviation). It is preferred to use matrix correlation instead of matrix covariance when there are differences on the scale of the variables; but, as our variables are on a Likert scale of 1,2,3, we are going to use correlation one, we think is more appropriate. Then, we are going to extract based on eigenvalues greater than 1. In rotation nothing, but we are going to ask for load chart, to show us the variables plane about the principal components. Related to scores, we are going to save those as variables, to get the score of each variable about the principal components, and to see ponderations of the components.

Advertencias

Sólo se ha extraído un componente. No se pueden generar gráficos de componente.

Matriz de correlaciones^a

| | | Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | Are women better or worse: on earnings earned for the same job? | Are women better or worse: on career opportunities / promotion? |
|-------------|---|---|---|---|
| Correlación | Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | 1,000 | ,554 | ,517 |
| | Are women better or worse: on earnings earned for the same job? | ,554 | 1,000 | ,498 |
| | Are women better or worse: on career opportunities / promotion? | ,517 | ,498 | 1,000 |
| | Are women better or worse: on job stability? | ,528 | ,524 | ,528 |
| | Are women better or worse: on adequate consideration by the employer and co-workers? | ,398 | ,377 | ,478 |

Matriz de correlaciones^a

| | | Are women better or worse: on job stability? | Are women better or worse: on adequate consideration by the employer and co-workers? |
|-------------|---|--|--|
| Correlación | Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | ,528 | ,398 |
| | Are women better or worse: on earnings earned for the same job? | ,524 | ,377 |
| | Are women better or worse: on career opportunities / promotion? | ,528 | ,478 |
| | Are women better or worse: on job stability? | 1,000 | ,491 |
| | Are women better or worse: on adequate consideration by the employer and co-workers? | ,491 | 1,000 |

a. Determinante = ,188

As we see, it extracts ONE principal component, and it generates a column named FAC1_1 with the scores. But with only one principal component we cannot generate components graphics.

As the determinant is not near 0 (it is 0.188), that means the correlation between variables is not so strong, so the PCA lose sense.

Comunalidades

| | Inicial | Extracción |
|---|---------|------------|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | 1,000 | ,614 |
| Are women better or worse: on earnings earned for the same job? | 1,000 | ,593 |
| Are women better or worse: on career opportunities / promotion? | 1,000 | ,622 |
| Are women better or worse: on job stability? | 1,000 | ,646 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | 1,000 | ,490 |

Método de extracción: análisis de componentes principales.

It this graphic above, we can analyze that V4: Are women better or worse: on job stability? Can be explain in a 64.6% by the principal component extracted. We must be careful because only 49% of V5: Are women better or worse: on adequate consideration by the employer and co-workers? is explained. We have two options in that case: add another variable that explain better V5 or eliminate V5 from the analysis because it not strongly correlated with the rest of variables.

We are going to do the first option, we are going to generate another principal component, so we will have 2.

On SPSS, extraction, in order to chose eigenvalues higher than 1, we are going to fix the number of extracted factors as 2.

Comunalidades

| | Inicial | Extracción |
|---|---------|------------|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | 1,000 | ,715 |
| Are women better or worse: on earnings earned for the same job? | 1,000 | ,742 |
| Are women better or worse: on career opportunities / promotion? | 1,000 | ,627 |

| | | |
|--|-------|------|
| Are women better or worse: on job stability? | 1,000 | ,648 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | 1,000 | ,899 |

Método de extracción: análisis de componentes principales.

As we can see, now V5 is the best one the principal components can explain with 89.9%.

We come back with the other case, one component.

Varianza total explicada

| Componente | Total | Autovalores iniciales | | Sumas de cargas al cuadrado de la extracción | |
|------------|-------|-----------------------|-------------|--|---------------|
| | | % de varianza | % acumulado | Total | % de varianza |
| 1 | 2,963 | 59,260 | 59,260 | 2,963 | 59,260 |
| 2 | ,669 | 13,376 | 72,635 | | |
| 3 | ,480 | 9,590 | 82,226 | | |
| 4 | ,446 | 8,927 | 91,153 | | |
| 5 | ,442 | 8,847 | 100,000 | | |

Varianza total explicada

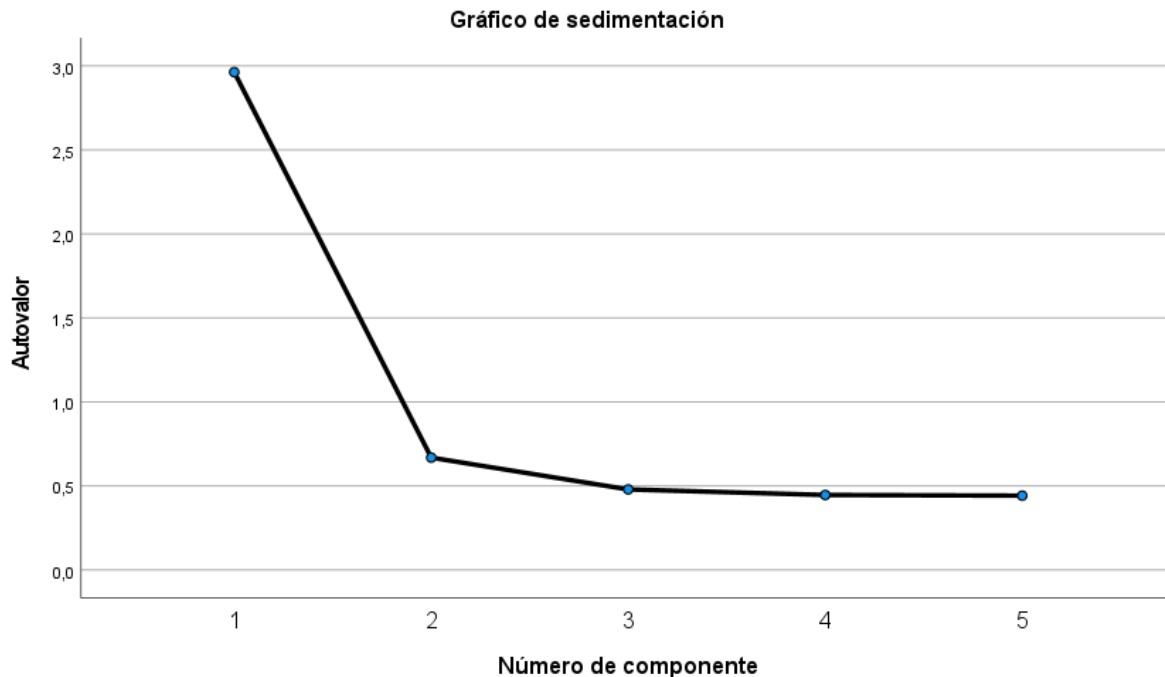
Sumas de cargas
al cuadrado de la
extracción

| Componente | % acumulado |
|------------|-------------|
| 1 | 59,260 |
| 2 | |
| 3 | |
| 4 | |
| 5 | |

Método de extracción: análisis de componentes principales.

This table above (Total variance explained) shows us that the total variance sums 5, as we have 5 variables, and the eigenvalues are ordered in a descendent way. The component eigenvalue higher than 1 means that this explains more than a variable itself.

$2.963/5 = 59.260\%$, the component 1 is explaining 59.260% of the information of the dataset.



Here we have the scree plot. The “elbow” corresponds to the point after which the eigenvalues decrease more slowly, which shows we can have 1 and even 2 principal components, because after that the graph is flattered.

Matriz de componente^a

| | Componente 1 |
|---|-----------------|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | ,783 |
| Are women better or worse: on earnings earned for the same job? | ,770 |
| Are women better or worse: on career opportunities / promotion? | ,788 |
| Are women better or worse: on job stability? | ,803 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | ,700 |

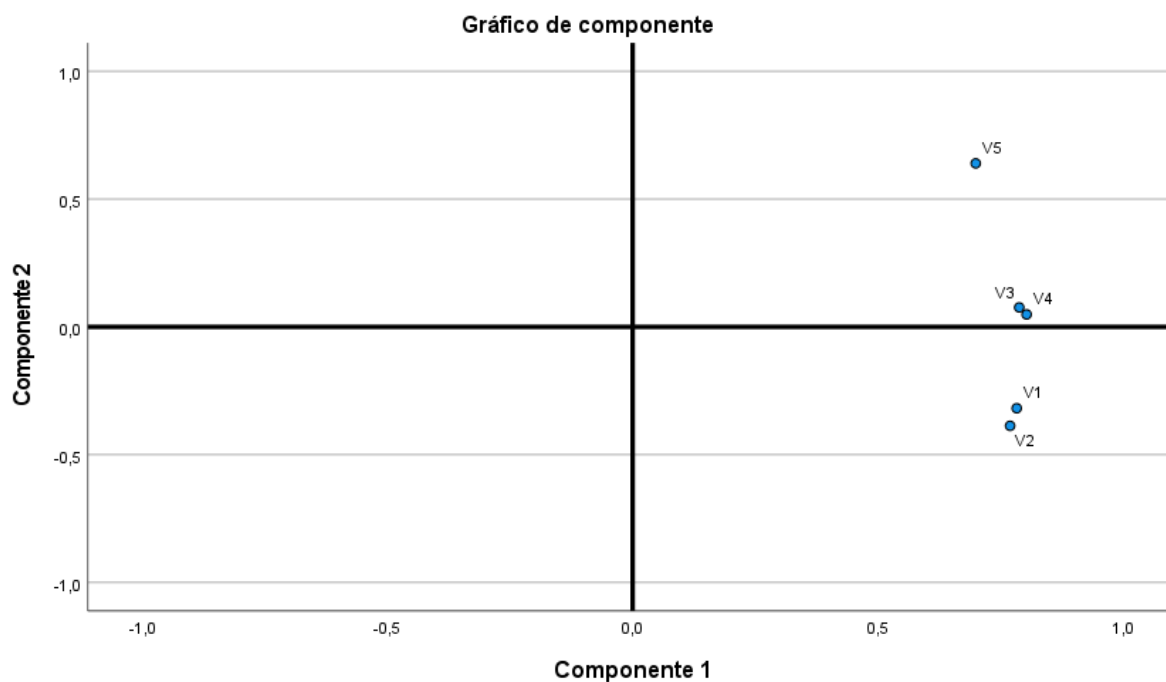
Método de extracción: análisis de componentes principales.^a

a. 1 componentes extraídos.

Component matrix say that component 1 explains better V4, and the worse V5; but all grow in the same direction.

We can observe better this in the components graphic with 2 components we are talking before.

So we can see how V1, V2 component 2 go in a negative direction, and they are in other quadrant.



Matriz de componentes

| | Componente | |
|---|------------|-------|
| | 1 | 2 |
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | ,783 | -,318 |
| Are women better or worse: on earnings earned for the same job? | ,770 | -,387 |
| Are women better or worse: on career opportunities / promotion? | ,788 | ,076 |
| Are women better or worse: on job stability? | ,803 | ,049 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | ,700 | ,640 |

Método de extracción: análisis de componentes principales.^a

a. 2 componentes extraídos.

Matriz de coeficiente de puntuación de componente

| | Componente 1 |
|---|-----------------|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | ,264 |
| Are women better or worse: on earnings earned for the same job? | ,260 |
| Are women better or worse: on career opportunities / promotion? | ,266 |
| Are women better or worse: on job stability? | ,271 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | ,236 |

Método de extracción: análisis de componentes principales.

Puntuaciones de componente.

Here we have the punctuations of the components. If we multiply 0.264 for the first row V1 standardize plus 0.260 first row V2 standardize...= first row of FAC1_1. Happen the same with more components.

Matriz de coeficiente de puntuación de componente

| | Componente 1 | 2 |
|---|-----------------|-------|
| Are women better or worse: possibility of finding a job suitable for their educational qualifications and / or experience | ,264 | -,476 |
| Are women better or worse: on earnings earned for the same job? | ,260 | -,579 |
| Are women better or worse: on career opportunities / promotion? | ,266 | ,114 |
| Are women better or worse: on job stability? | ,271 | ,073 |
| Are women better or worse: on adequate consideration by the employer and co-workers? | ,236 | ,957 |

Método de extracción: análisis de componentes principales.

Puntuaciones de componente.

Varianza total explicada

| Componente | Total | Autovalores iniciales | | Sumas de cargas al cuadrado de la extracción | |
|------------|-------|-----------------------|-------------|--|---------------|
| | | % de varianza | % acumulado | Total | % de varianza |
| 1 | 2,963 | 59,260 | 59,260 | 2,963 | 59,260 |
| 2 | ,669 | 13,376 | 72,635 | ,669 | 13,376 |
| 3 | ,480 | 9,590 | 82,226 | | |
| 4 | ,446 | 8,927 | 91,153 | | |
| 5 | ,442 | 8,847 | 100,000 | | |

The last thing to check is that Components 1 and 2 are uncorrelated, so we make a correlation test and obtain 0.

Correlaciones

| | | REGR factor score 1 for analysis 3 | REGR factor score 2 for analysis 3 |
|------------------------------------|------------------------|------------------------------------|------------------------------------|
| REGR factor score 1 for analysis 3 | Correlación de Pearson | 1 | ,000 |
| | Sig. (bilateral) | | 1,000 |
| | N | 304 | 304 |
| REGR factor score 2 for analysis 3 | Correlación de Pearson | ,000 | 1 |
| | Sig. (bilateral) | 1,000 | |
| | N | 304 | 304 |

7) Using the Regression Tree, identify the tree using the First component of PCA (previous point) as the dependent variable and GENDER, AGE, PNAS, EDUC, CONDIZ, NCOMP as independent variables. Describe the results.

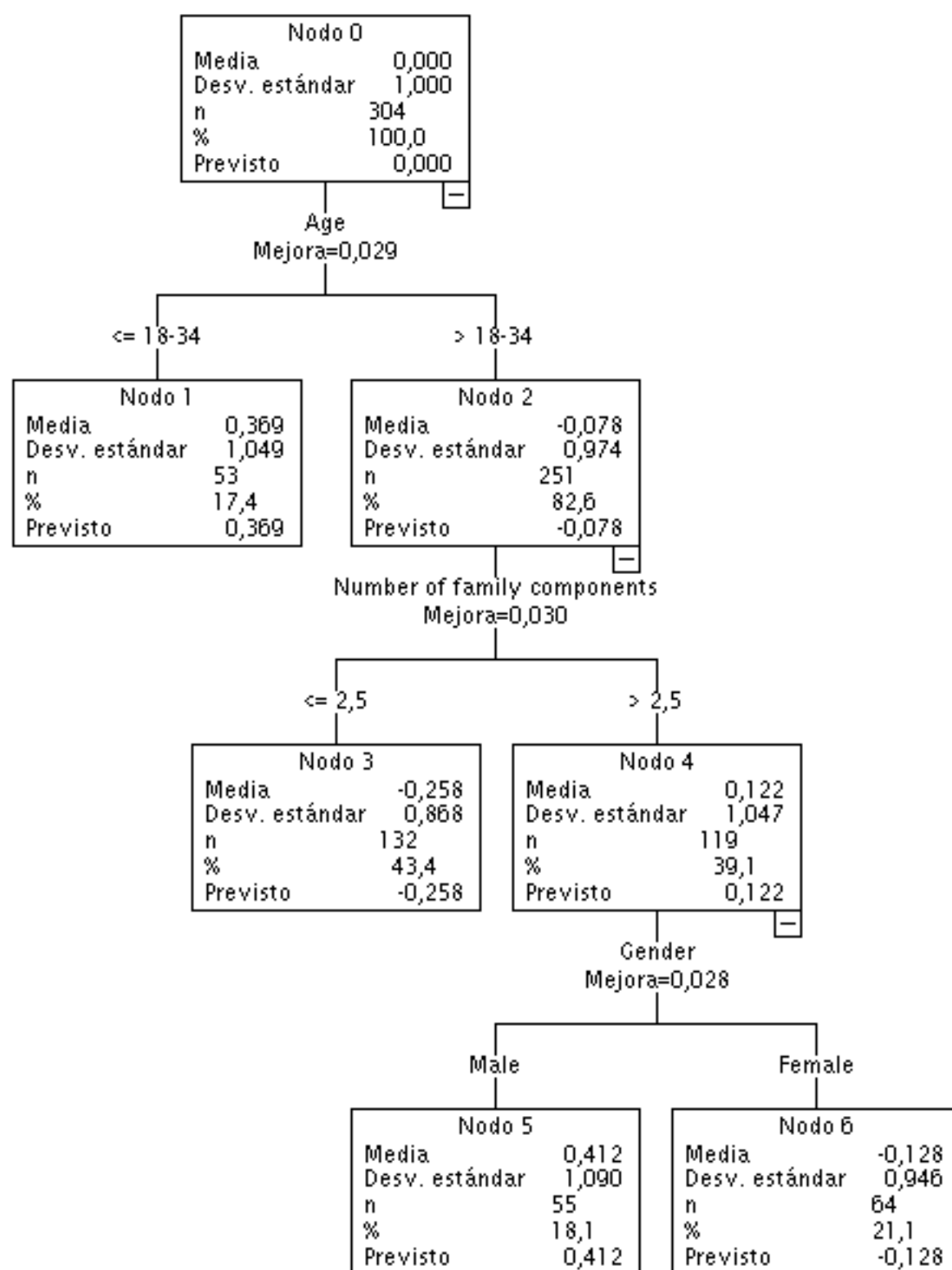
In SPSS, we must go to analyze, classify, and tree.

The growing method we are using is CRT; the First component of PCA (previous point) as the dependent variable and GENDER, AGE, PNAS, EDUC, CONDIZ, NCOMP as independent variables; the statistics outputs and plots must be importance to model; the criteria we have selected this parent nodes: 50 for child nodes and 100 for parent nodes because we have a considerable length of observations in the dataset; and we are not pruning the tree to avoid overfitting.

Resumen del modelo

| | | |
|------------------|------------------------------------|--|
| Especificaciones | Método de crecimiento | CRT |
| | Variable dependiente | REGR factor score 1 for analysis 1 |
| | Variables independientes | Gender, Age, Country of birth, Level of Education, Work condition, Number of family components |
| | Validación | Ninguna |
| | Máxima profundidad del árbol | 5 |
| | Casos mínimos en nodo padre | 100 |
| | Casos mínimos en nodo hijo | 50 |
| Resultados | Variables independientes incluidas | Age, Work condition, Number of family components, Level of Education, Gender |
| | Número de nodos | 7 |
| | Número de nodos terminales | 4 |
| | Profundidad | 3 |

REGR factor score 1 for analysis 1



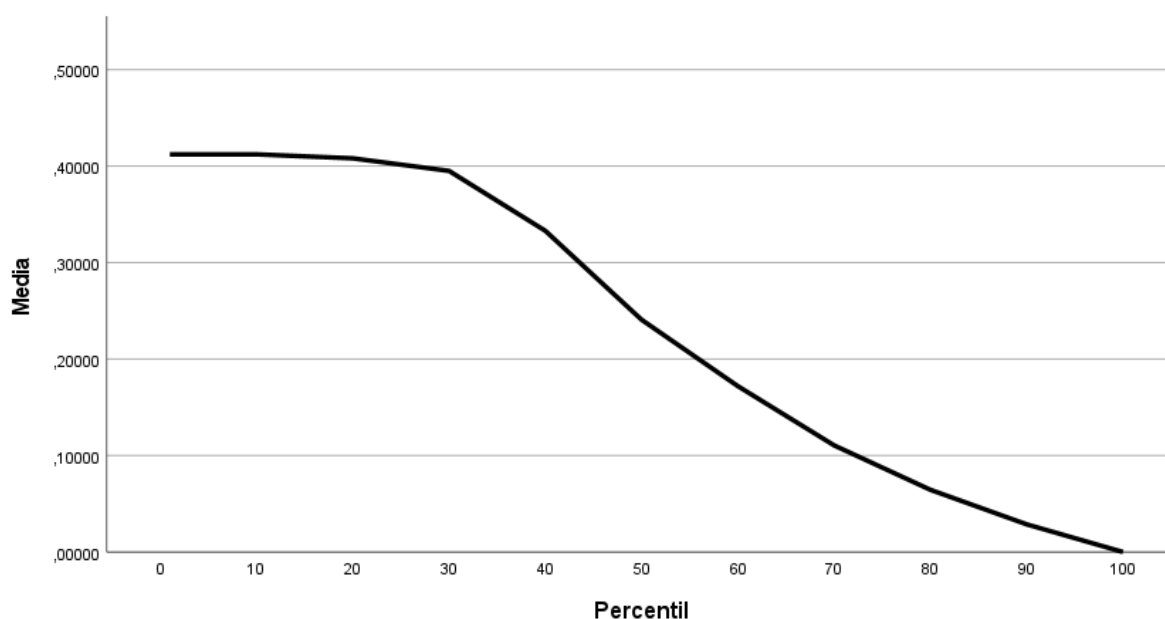
Resumen de ganancias para nodos

| Nodo | N | Porcentaje | Media |
|------|-----|------------|-----------|
| 5 | 55 | 18,1% | ,4122844 |
| 1 | 53 | 17,4% | ,3690147 |
| 6 | 64 | 21,1% | -,1279725 |
| 3 | 132 | 43,4% | -,2579029 |

Método de crecimiento: CRT

Variable dependiente: REGR factor score 1 for analysis 1

Here we can see that are 4 terminal nodes, these are the nodes without children.



Método de crecimiento: CRT

Variable dependiente: REGR factor score 1 for analysis 1

Riesgo

| Estimación | Error estándar |
|------------|----------------|
| ,910 | ,081 |

Método de crecimiento: CRT

Variable dependiente: REGR factor score 1 for analysis 1

Importancia de variable independiente

| Variable independiente | Importancia | Importancia normalizada |
|-----------------------------|-------------|-------------------------|
| Age | ,034 | 100,0% |
| Number of family components | ,030 | 86,5% |
| Gender | ,028 | 82,7% |
| Work condition | ,021 | 61,3% |
| Level of Education | ,020 | 59,2% |

Método de crecimiento: CRT
Variable dependiente: REGR factor score 1 for analysis 1

